Bayes factors for linear regression

Rui Wang

Tuesday 9^{th} October, 2018

1 Introduction

This note gives a review for Bayes factors for linear regression.

2 Mixture of g prior

This section is adapted from Liang et al. (2008). Suppose $\mathbf{Y} \in \mathbb{R}^n$ is generated from the model

$$\mathcal{M}_\gamma : \mathbf{Y} = \mathbf{1}_n \alpha + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

where $\mathbf{X} \in \mathbb{R}^{n\times p}$ and $\boldsymbol{\varepsilon} \sim \mathcal{N}(0, \phi^{-1}\mathbf{I}_n)$.

Let $\mathbf{X}_\gamma \in \mathbb{R}^{n\times p_\gamma}$ be a submatrix of $\mathbf{X}$. Then the submodel $\mathcal{M}_\gamma$ is defined as

$$\mathcal{M}_\gamma : \mathbf{Y} = \mathbf{1}_n \alpha + \mathbf{X}_\gamma\boldsymbol{\beta}_\gamma + \boldsymbol{\varepsilon}.$$

The null model $\mathcal{M}_N$ is

$$\mathcal{M}_\gamma : \mathbf{Y} = \mathbf{1}_n \alpha + \boldsymbol{\varepsilon}.$$

We would like to compare $\mathcal{M}_\gamma$ with $\mathcal{M}_N$. Without loss of generality, we assume $\mathbf{1}_n^\top \mathbf{X}_\gamma = 0$. Under $\mathcal{M}_N$, the g prior is

$$p(\alpha, \phi | \mathcal{M}_N) = \frac{1}{\phi}.$$

Under $\mathcal{M}_\gamma$, the g prior is

$$\boldsymbol{\beta}_\gamma | \phi, \mathcal{M}_\gamma \sim \mathcal{N}(0, \frac{g}{\phi}(\mathbf{X}_\gamma^\top \mathbf{X}_\gamma)^{-1}), \quad p(\alpha|\phi, \mathcal{M}_\gamma) \propto 1, \quad p(\phi|\mathcal{M}_\gamma) = \frac{1}{\phi}.$$

The joint pdf is

$$\begin{aligned}
&p(\mathbf{Y}, \alpha, \boldsymbol{\beta}_\gamma, \phi | \mathcal{M}_\gamma) = p(\mathbf{Y}|\alpha, \boldsymbol{\beta}_\gamma, \phi, \mathcal{M}_\gamma) p(\boldsymbol{\beta}_\gamma|\phi, \mathcal{M}_\gamma) p(\alpha|\phi, \mathcal{M}_\gamma) p(\phi|\mathcal{M}_\gamma) \\
=&(2\pi)^{-(n+p_\gamma)/2} g^{-p_\gamma/2} \phi^{(n+p_\gamma)/2-1} |\mathbf{X}_\gamma^\top \mathbf{X}_\gamma|^{1/2} \exp\left\{-\frac{n\phi}{2}(\bar{\mathbf{Y}} - \alpha)^2\right\} \\
&\exp\left\{-\frac{\phi(g+1)}{2g}\left\|\mathbf{X}_\gamma\left(\boldsymbol{\beta}_\gamma - \frac{g}{g+1}\hat{\boldsymbol{\beta}}_\gamma\right)\right\|^2 - \frac{\phi}{2(g+1)}\left\|\mathbf{X}_\gamma\hat{\boldsymbol{\beta}}_\gamma\right\|^2 - \frac{\phi}{2}\left\|\mathbf{Y} - \mathbf{1}_n\bar{\mathbf{Y}} - \mathbf{X}_\gamma\hat{\boldsymbol{\beta}}_\gamma\right\|^2\right\},
\end{aligned}$$

where $\bar{\mathbf{Y}} = n^{-1}\mathbf{1}_n^\top \mathbf{Y}$, $\hat{\boldsymbol{\beta}}_\gamma = (\mathbf{X}_\gamma^\top \mathbf{X}_\gamma)^{-1}\mathbf{X}_\gamma^\top \mathbf{Y}$.

Direct calculation yields

$$p(\mathbf{Y}|\mathcal{M}_\gamma, g) = \frac{\Gamma((n-1)/2)}{\pi^{(n-1)/2}\sqrt{n}} \left\|\mathbf{Y} - \mathbf{1}_n\bar{\mathbf{Y}}\right\|^{-(n-1)} \frac{(1+g)^{(n-p_\gamma-1)/2}}{[1+g(1-R_\gamma^2)]^{(n-1)/2}},$$

where $R_\gamma^2 = 1 - \|\mathbf{Y} - \mathbf{1}_n\bar{\mathbf{Y}} - \mathbf{X}_\gamma\hat{\boldsymbol{\beta}}_\gamma\|^2/\|\mathbf{Y} - \mathbf{1}_n\bar{\mathbf{Y}}\|^2$. Also, we have

$$p(\mathbf{Y}|\mathcal{M}_N) = \frac{\Gamma((n-1)/2)}{\pi^{(n-1)/2}\sqrt{n}} \left\|\mathbf{Y} - \mathbf{1}_n\bar{\mathbf{Y}}\right\|^{-(n-1)}.$$

Thus,

$$\mathrm{BF}[\mathcal{M}_\gamma : \mathcal{M}_N] = (1+g)^{(n-p_\gamma-1)/2}[1+g(1-R_\gamma^2)]^{-(n-1)/2}.$$

2.1 Choices of g

Local empirical Bayes. The local EB estimates a separate g for each model $\mathcal{M}_\gamma$.

$$\hat{g}_\gamma^{\mathrm{EBL}} = \arg\max_{g\geq 0} p(\mathbf{Y}|\mathcal{M}_\gamma, g) = \arg\max_{g\geq 0} \frac{(1+g)^{(n-p_\gamma-1)/2}}{[1+g(1-R_\gamma^2)]^{(n-1)/2}} = \max\{F_\gamma - 1, 0\},$$

where

$$F_\gamma = \frac{R_\gamma^2/p_\gamma}{(1-R_\gamma^2)/(n-1-p_\gamma)}$$

is the usual F statistic for testing $\boldsymbol{\beta}_\gamma = 0$.

Global empirical Bayes. The global EB procedure assumes one common g for all models.

$$\hat{g}_\gamma^{\mathrm{EBG}} = \arg\max_{g\geq 0} \sum_\gamma p(\mathcal{M}_\gamma) p(\mathbf{Y}|\mathcal{M}_\gamma, g) = \arg\max_{g\geq 0} \sum_\gamma p(\mathcal{M}_\gamma) \frac{(1+g)^{(n-p_\gamma-1)/2}}{[1+g(1-R_\gamma^2)]^{(n-1)/2}}.$$

References

Liang, F., Paulo, R., Molina, G., Clyde, M. A., and Berger, J. O. (2008). Mixtures of g priors for bayesian variable selection. *Journal of the American Statistical Association*, 103(481):410–423.